Towards Computational Neuroinformatics: Information Theoretic Analyses of Neural Coding

David H. Goldberg[1], Daniel Gardner[1], and Jonathan D. Victor[2]

[1]Lab of Neuroinformatics, Dept. of Physiology, [2]Dept. of Neurology & Neuroscience, Weill Cornell Med. Coll., NYC, NY

Towards Computational Neuroinformatics

Initiating the second decade of the Human Brain Project, we announce a new collaborative thrust to synthesize computational neuroscience and neuroinformatics. New techniques of computational neuroinformatics will be developed and brought to bear on the fundamental problem of neural coding—how neurons represent and process information.

Current analyses and models are often tested on datasets from one preparation or protocol, limiting the ability to test theories and distinguish general from domain-specific. To remove limits due to restricted exchange of algorithms or datasets, we are enabling a wide range of project-developed and user-submitted parallelized computational algorithms to be applied to a correspondingly broad assemblage of neural datasets collected from multiple cortical areas, many protocols, and several preparations. The project synthesizes:

- a suite of parallelized data-driven information-based analytical algorithms, with
- a powerful multiprocessor computational array, each linked to
- open, public databases that acquire, aggregate, and archive single- and multi-unit neuronal data from visual and somatosensory cortex, and other vertebrate and invertebrate networks collected by multiple labs, techniques, data formats and protocols.

A Collaboration of Computational Neuroscience and Neuroinformatics

- Project informatics and neuroscience collaborators and consultants include: A.B. Bonds, E. N. Brown, P. M. DiLorenzo, R. Elber, E. P. Gardner, P. Mitra, S. Nirenberg, K. Purpura, R. C. Reid, P. Reinagel, B. J. Richmond, S. Schultz, R. Zabih; many others have designed algorithms.
- We actively invite additional collaborators—to develop, supply or test algorithms, to provide experimental neuroscience data, or to explore parallelization—for this effort to advance our understanding of neural coding. We also seek to advance the design and execution of protocols for the study of visual and somatic sensation that focus on hypothesis testing using these evolving algorithms.
- The expanding Laboratory of Neuroinformatics seeks biophysicists and neuroscientists with additional expertise in computer science, computational methods, software development, bioinformatics, or information theory.
- Contact one of us or any of our LNI colleagues at the meeting, or e-mail dan@aplysia.med.cornell.edu

Our Neuroinformatics Resources:

- Neurodatabase & tools: neurodatabase.org
- BrainML and BrainMetaL: brainml.org
- Data sharing policies: datasharing.net

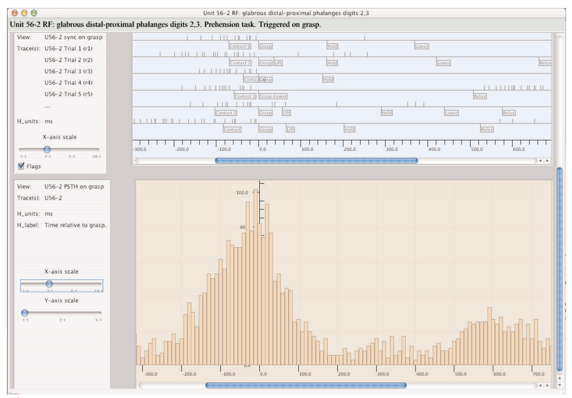
Acknowledgments:

Human Brain Project / Neuroinformatics research funded via MH68012 from NIMH, NINDS, NIA, NIBIB & NSF, and MH57153 from NIMH & NINDS.

LNI Laboratory of Neuroinformatics
Weill Medical College of Cornell University

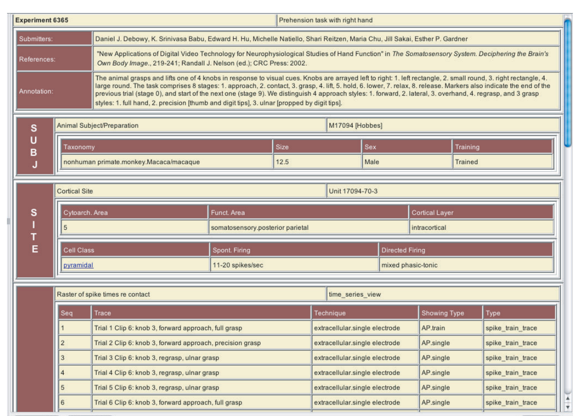
Computational Neuroinformatics Extends Existing Neurodatabase.org

Accessible via neurodatabase.org, our Human Brain Project supported, open, public neurodatabase currently allows users to archive and download actual datasets and descriptive metadata from a variety of preparations, including mammalian cortex. Java-based tools support multiplatform assembly, upload, annotation, search, and acquisition, as well as viewing of XML-wrapped data on expanded or contracted time scales, accompanied by descriptive metadata (Figs. 1 & 2, *below*).



1. VirtualOscilloscope Displays Databased Traces

The VirtualOscilloscope multiplatform Java tool provides figure-like display of datasets archived at neurodatabase.org. Variable sweep and gain transcends limitations of the printed page and enhances examination and re-use.



2. Descriptive Metadata Aid Interpretation and Analysis of Shared Neurophysiological Data

Re-use of neurophysiological datasets requires descriptive metadata that specify data origin and techniques. Such controlled-vocabulary metadata from neurodatabase.org are displayed here. Metadata are grouped into intuitive categories covering submitters, journal reference, site of recording, and the datasets themselves, organized as multiple traces in a view, each specifying technique, data type, and display class. Buttons below allow export as CSV or XML, or plotting via the Virtual Oscilloscope.

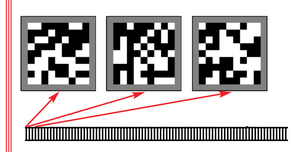
Information Theory Can Inform Neural Coding

The spiking activity of populations of neurons is the basis for the brain's representation and processing of sensations, cognitive states, and actions. However, the roles of various aspects of neural activity (firing rate, spike times, synchrony, oscillations, spatial pattern) are imperfectly understood at present. Information theory provides a sound theoretical basis for addressing these issues, and measuring information from neural data can give insight into the contributions of various features of neural activity to the above processes.

Multiple Analytical Methods Can Provide Complementary Insights

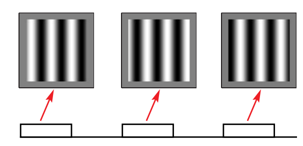
Different methods of information calculation have complementary domains of applicability and limitations (Fig. 5). Applying multiple analytical methods to different stimulus paradigms provides complementary insights to information coding.

As a case study, we ask how information is coded in local clusters of neurons, analyzing data from two studies of neuron clusters in macaque V1 simultaneously recorded via tetrodes.



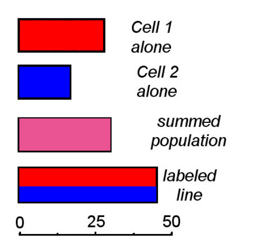
3a. Continuous Stimuli

Visual stimuli are continuously-modulated pseudorandom checkerboards at 14 ms/frame.



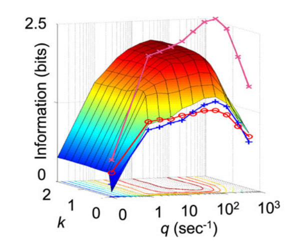
3b. Sequential Stimuli

Visual stimuli are sequentially spaced gratings in 16 spatial phases, single spatial frequency and orientation at 250ms/frame.



4a. Analysis Using Direct Method

- Information conveyed by neuron of origin eliminates redundancy
- Results extend to clusters of up to 6 neurons
- Results are restricted to information rates over brief times (8 to 15 ms)



Aronov et al 2003; as for Fig. 4a, blue and red curves show individual neurons; magenta shows summed data.

4b. Analysis Using Metric-Space Method

- Information conveyed by neuron of origin reduces but does not eliminate redundancy
- Results not yet extended beyond pairs of neurons
- Results apply to extended periods of time (up to 500 ms)
- Analysis method characterizes the temporal precision of coding

Collaborative Development and Implementation of Multiple Analytic Algorithms and Their Application to the Question of Neural Coding

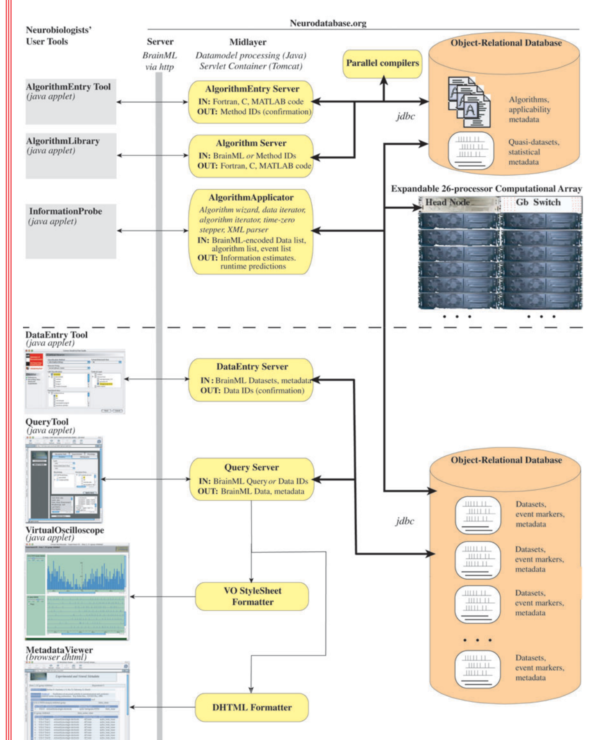
We are building an open, collaborative resource to implement and apply parallelized analytic algorithms to to a database of neural data from multiple areas, protocols, and preparations. Such algorithms will quantify the information present in patterns of neural activity and determine which features of activity carry information, transmit it between neurons and areas, and relate it to behavior.

Neuronal signaling is plausibly system- and context-dependent, so that different systems and regions have evolved partially-distinct biological solutions to the coding problem. Identifying a particular coding strategy thus requires testing with a broad set of approaches. We are therefore implementing multiple analytical algorithms that differ along many axes, including suitability for multichannel data, suitability for continuous signals as well as spike trains, and the extent of the assumptions concerning the nature of neural coding.

Method	Data Types			Neural Coding			Parallelism
	Stimulus applicability	Suitability for multichannel/ multiunit data*	Suitable for continuous signals	Assumptions concerning nature of code#	Analyzes response space	Provides a description of the code	Suitability for parallel processing
direct	general	low	no	virtually none	no	no	yes
reconstruction	time series	favorable	yes	Volterra	no	yes	yes
metric space	general	very favorable	no	specific models	yes	yes	yes
embedding	general	favorable	yes	continuous	yes	minimal	yes
power series	general	favorable	no	Volterra	no	yes	yes
spectrotemporal	general	favorable	yes	continuous	no	yes	yes
bottleneck/codebook	general	very favorable	yes	continuous	yes	yes	yes

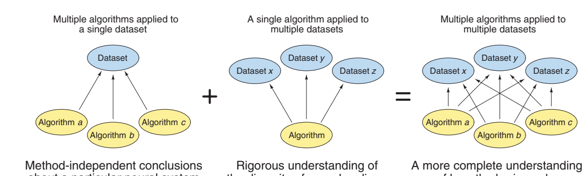
5. Information-Theoretic Analytical Methods to be Implemented at Neurodatabase.org

key: * *low:* exponential growth with number of channels
* *favorable:* polynomial growth (for low-order models)
* *very favorable:* no growth
\# *from most to least general:* virtually none, continuous, Volterra, specific models



6. Client-Server System for Data and Algorithms

Upper: in-development Java tools, midlayer servers, algorithm database and computational cluster. *Lower:* underlying tools, server, and object-relational database for spike trains and other neurophysiological datasets at neurodatabase.org



7. Towards an Understanding of Neural Coding

By examining a given dataset using multiple approaches (*left*), or by applying a selected approach to multiple datasets (*center*), investigators can gain confidence that neurobiological conclusions reflect broad principles. Our in-development resources will also allow multiple algorithms to be applied to multiple datasets (*right*) to reveal both general principles and specific classes of neural coding.

A Community Resource for Exploring Neural Coding

These extensions to neurodatabase.org are thus designed to:

- implement multiple algorithms for extracting features of neural coding,
- serve data for post-hoc examination and analyses,
- serve algorithms for local analyses of both local and global data,
- link to a parallel open archive for data,
- provide an open archive for algorithms,
- establish a site and facilitate procedures for both collaborative and open development and evolution of algorithms,
- offer a computational engine for testing and application of single or multiple algorithms to specific datasets or multiple groups of datasets, and
- publish turnkey specs aiding local implementation of algorithmic and computational resources.